Fake News Detection Based on the Correlation Extension of Multimodal Information

Yanqiang Li[1,2], Ke Ji[1,2(✉)], Kun Ma[1,2], Zhenxiang Chen[1,2], Jin Zhou[1,2], and Jun Wu[3]

[1] School of Information Science and Engineering, University of Jinan, Jinan 250022, China
ise_jik@ujn.edu.cn
[2] Shandong Provincial Key Laboratory of Network Based Intelligent Computing, University of Jinan, Jinan 250022, China
[3] School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

Abstract. Online social media is characterized by a large number of users that creates conditions for large-scale news generation. News in multimodal form (images and text) often has a serious negative impact. Existing multimodal fake news detection methods mainly explore the relationship between images and texts by extracting image features and text features. However, these methods typically ignore textual content in images and fail to explore the relationship between news and image texts further. We propose a new fake news detection method based on correlation extension multimodal (CEMM) information to solve this problem. The correlation between multimodal information is extended and the relationship between the extended image information and the news text is explored further by extracting text and statistical features from the image. This CEMM-based detection method consists of five parts, which can discover the relevant parts of news and optical character recognition (OCR) text and the features of fake news images and relevant parts of news text, and combine the information of the news itself to detect fake news. Experimental results proved the effectiveness of our approach.

Keywords: Fake news detection · Multimodal fusion · Deep learning

1 Introduction

News dissemination has gradually changed from the use of traditional electronic equipment, such as TV and radio, to the Internet, such as online social media. Compared with traditional electronic devices, news in the Internet environment spreads faster and has a wider impact, and the authenticity of the news cannot be guaranteed. In this case, fake news is a serious threat that will likely demonstrate negative effects.

Many people have been affected by fake news and suffered emotional or other types of harm or loss [1]. News with pictures is reposted 11 times more than

B. Li et al. (Eds.): APWeb-WAIM 2022, LNCS 13421, pp. 443–450, 2023.
https://doi.org/10.1007/978-3-031-25158-0_36

the average number of reposts of text-only news [6], thereby indicating that multimodal news can cause greater harm. The importance of multimodal forms of fake news detection has been demonstrated [5]. Therefore, exploring and solving the problem of detecting multimodal forms of fake news are necessary. Many approaches [13,14] have been proposed to solve this problem from a multimodal perspective. However, most of them use pre-trained models to extract image features, but ignore the text content on the image and the relationship between the text on the image and the news text.

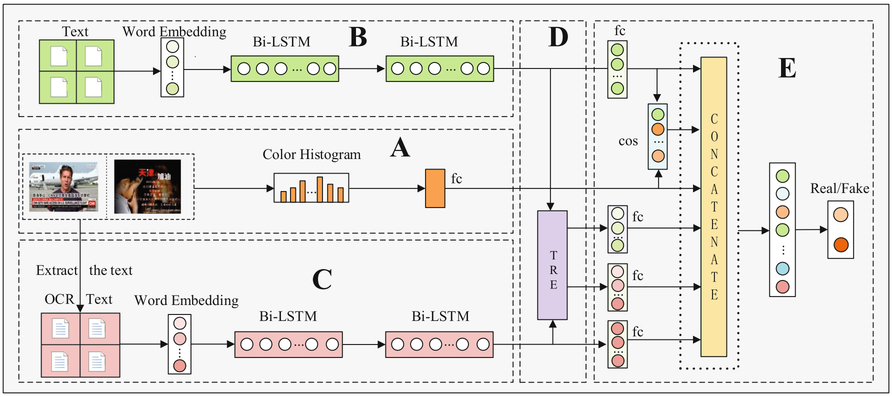


Fig. 1. Overview of our approach.

We propose a multimodal fake news detection model based on the optical character recognition (OCR) text in this study. Our model is shown in Fig. 1 and consists of five parts: (A) visual feature extractor, (B) news text feature extractor, (C) OCR text feature extractor, (D) text correlation extractor (TRE), and (E) feature integration classifier. Our method is significantly better than other approaches and validated using existing datasets.

We introduce the structure of the rest of the paper: Sect. 2 describes related work on fake news detection and text and images, Sect. 3 describes our approach, Sect. 4 presents the dataset, baseline, parameter settings, and comparison results, and Sect. 5 concludes the paper.

2 Related Work

2.1 Image Analysis and Text Classification

Image analysis is the description of the image content and the process of transforming the image described in pixels into data form by means of segmentation or feature extraction. Image analysis has used artificial intelligence or pattern recognition methods in recent years to describe, classify, and interpret images on the basis of image content. The text classification problem has formed the two subproblems of feature engineering and classifier with the development of machine learning. With the development of deep learning, deep learning related methods [4,9] have replaced traditional methods in feature engineering and classifier selection.

2.2 Fake News Detection

The detection was initially based on the text content of news. The content of pictures plays an important role in determining the authenticity of news. Therefore, detecting fake news from the perspective of images is challenging. An image is analyzed from the perspective of an event, and fake news was detected by extracting visual and statistical features of the image in the event [6]. Quality features and semantic information are extracted from the images, and the two features are fused using an attention mechanism to detect fake news images [11]. The attention mechanism is utilized to fuse image and text features to detect the authenticity of news [2,7,12]. A variational autoencoder is utilized to obtain a shared representation of text and images and utilize it to detect fake news [8]. The similarity between text and image is calculated, and features of text and image are added as the basis for fake news detection [17].

The utilization of image information (OCR text and statistical features) is insufficient in previous work, we further explore the relationship between image information and news text.

3 Our Approach

3.1 Problem Definition

Ω represents the dataset, and ω_k represents a sample in the dataset. $y_k \in (0, 1)$, where y_k is the label corresponding to the sample ω_k, 0 represents the true news, and 1 represents the fake news. Particularly, ω_k is composed of text and images.

3.2 Visual Feature Extractor

Images attached to fake news are typically visually impactful [11], and color is an intuitive indication of visual impact. The color histogram is a statistical feature of images for spam classification tasks [15]. This shows that color histogram features are effective in determining fake news images, therefore, we use color histogram and obtain the feature representation V of the image through a 32-dimensional fully connected layer.

3.3 News Text Feature Extractor

The news text T_{news} typically consists of words t^i_{news}. The pretrained language model (BERT) has been successfully used in many tasks. BERT can learn contextual word embedding and demonstrates enhanced generalization capabilities for other tasks after pretraining a large number of corpora. Therefore, we use the pretrained model BERT as our word embedding framework in part B of Fig. 1. We first determine the word embedding w_i that corresponds to the word. And we used two Bi-LSTMs to model the news text to obtain each word and its relationship n_i.

3.4 OCR Text Feature Extractor

We use a feature on the basis of images called OCR text feature (part C in Fig. 1). A given item of fake news includes text and image. The text on the image is crucial in the picture content and an important basis for judging whether the news is fake. A tesseract is an open-source OCR engine used for extracting the text from images and removing spaces and carriage return to obtain the OCR text. However, only some images contain text. We use 10 zero-padding techniques for images without text. And we use the same processing as news text to get OCR text features o_j.

3.5 Text Correlation Extractor

We obtain news and OCR text features extracted by the two layers of Bi-LSTM in the first two parts. We determine the relationship between OCR and news texts in part D of Fig. 1, that is, if news and OCR texts are related, then news and picture texts will likely describe the same thing. Therefore, we use soft alignment [3] on news and OCR texts to calculate the similarity matrix e_{ij} between them and then utilize it as the weight of the attention mechanism as follows [10]:

$$e_{ij} = n_i^{\mathrm{T}} o_j. \tag{1}$$

We use the method of [10] to determine the relationship between the two texts as follows:

$$\overline{n}_i = \sum_{j=1}^{l_o} \frac{exp(e_{ij})}{\sum_{k=1}^{l_o} exp(e_{ik})} o_j, \forall i \in [1_1, \cdots, l_n], \tag{2}$$

$$\overline{o}_j = \sum_{i=1}^{l_n} \frac{exp(e_{ij})}{\sum_{k=1}^{l_n} exp(e_{kj})} n_i, \forall j \in [1_1, \cdots, l_o]. \tag{3}$$

We intuitively find the word that is semantically similar to the news text from the OCR text using (2) and use $\overline{n}_i$ to indicate it. We first calculate the weight of elements in the i-th row and the j-th column in the i-th row in e_{ij} and multiply it with o_j, that is, multiply the weight of the element e_{ij} in e_{i*} with o_j. Finally, the above calculation process is performed for each element in e_{i*} and the sum is obtained to represent the degree of correlation between n_i (i-th word in the news text) and the OCR text. The other formula has the same meaning.

3.6 Feature Integration Classifier

We first summon previous visual, two text, and correlation features into features of the same dimension through the fully connected layer for fusion in part E of Fig. 1. We calculate the cosine similarity between V and n and express it with cos to determine the relationship between news text and images. This method has been effectively utilized in many tasks [16,17].

We then determine whether the news is true or false on the basis of existing information using the following formula:

$$Y = fnc([n_i, o_j, \overline{n}_i, \overline{o}_j, cos, V]), \tag{4}$$

where fnc is the fake news classifier, [ ] is the concatenate operation, and Y is the prediction label.

4 Experiments

4.1 Dataset

Table 1. Details of the dataset.

Category	News	Image	OCR
Fake news	19285	10768	3286
Real news	19186	11064	5276
Total	38471	21832	8562

The dataset[1] in Table 1 we used was obtained from the AI competition platform BIENdata and jointly published by the Institute of Computer Technology of the Chinese Academy of Sciences and Beijing Academy of Artificial Intelligence.

4.2 Baseline

Relevant experiments are performed on the existing dataset to verify the effectiveness of our method.

Bi-LSTM-Att: We use word2vec to convert news text into word embeddings and use Bi-LSTM and attention mechanism to classify news text.

Color Histogram: We employ the color histogram feature to classify using the BPNN model.

OCR Text: The OCR text is processed in the same way as part D in Fig. 1 to obtain the classification results.

att-RNN: We removed the social features in att-RNN [7] to get the classification results.

MVAE: We use MVAE [8] to get the classification results.

[1] https://www.biendata.xyz/competition/falsenews/

Table 2. Experimental results.

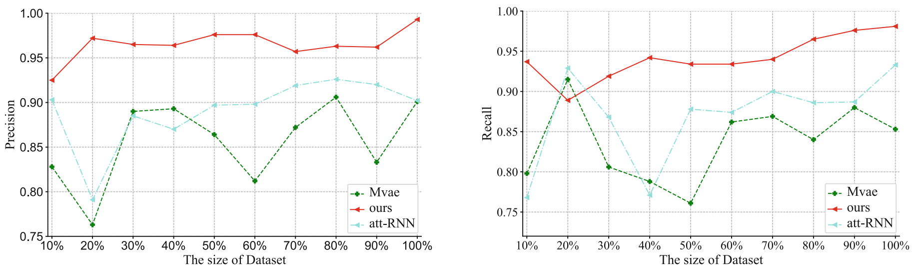
Category	Method	Accuracy	Fake news			Real news		
			Precision	Recall	F1	Precision	Recall	F1
News text	Bi-LSTM-Att	0.874	0.900	0.843	0.871	0.850	0.905	0.877
Visual	Color Histogram	0.639	0.615	0.759	0.679	0.679	0.517	0.587
OCR text	Bi-LSTMs	0.545	0.530	0.851	0.653	0.609	0.236	0.340
Multimodal	att-RNN	0.907	0.926	0.886	0.905	0.889	0.928	0.908
	MVAE	0.881	0.880	0.885	0.882	0.882	0.877	0.880
	CEMM	**0.964**	**0.974**	**0.954**	**0.964**	**0.955**	**0.974**	**0.964**

4.3 Parameter Setting

Bi-LSTM consists of two layers with a cell size of 32 and is utilized for processing text. The dimension of the color histogram is 1024. The classifier consists of two fully connected layers. One layer presents a cell size of 64, and the other layer contains a sigmoid activation function. The learning rate is set to 0.0002, and the optimizer uses Adam.

4.4 Comparison Results

Table 2 Results Analysis: The comparison of the experimental results demonstrated that the accuracy and precision of our model are better than those of other models. This finding indicated that our method performs better than other approaches in the task of identifying fake news. The performance of MVAE is similar to Bi-LSTM-Att because the proportion of news containing images in our dataset is only 57%, which may lead to the inability of MVAE to learn a multimodal shared representation.



(a) Precision of fake news (b) Recall of fake news

Fig. 2. Variations of different proportions of indicators in the training dataset.

Figure 2 shows the changes in evaluation indicators of different methods using various training data ratios. And Fig. 2 shows that the performance of our method is significantly higher than that of other models.

Table 3. Performance of the model after removing each part.

Method	Accuracy	Fake news			Real news		
		Precision	Recall	F1	Precision	Recall	F1
CEMM	0.964	0.974	0.954	0.964	0.955	0.974	0.964
CEMM^{-bert}	0.925	0.928	0.920	0.924	0.921	0.929	0.925
$\mathrm{CEMM}^{-bert\ tre}$	0.923	0.936	0.908	0.922	0.911	0.939	0.925
$\mathrm{CEMM}^{-bert\ ocr\ tre}$	0.914	0.960	0.863	0.909	0.876	0.964	0.918
$\mathrm{CEMM}^{-bert\ cos}$	0.923	0.943	0.899	0.921	0.904	0.946	0.925
$\mathrm{CEMM}^{-bert\ ch}$	0.922	0.923	0.920	0.922	0.920	0.924	0.922

Table 3 Experimental Settings: Table 3 shows the effect of removing certain parts of CEMM.

$CEMM^{-bert}$: We remove BERT on CEMM and use word2vec for word embedding.

$CEMM^{-bert\ tre}$: We remove the text relevance extractor (tre) on the basis of removing BERT.

$CEMM^{-bert\ ocr\ tre}$: We remove the OCR text and the text relevance extractor on the basis of removing BERT.

$CEMM^{-bert\ cos}$: We remove the cosine similarity on the basis of removing BERT.

$CEMM^{-bert\ ch}$: We remove the color histogram (ch) after removing BERT.

Table 3 Results Analysis: The performance of our method is better than other methods (att-RNN and MVAE) after replacing the pretrained BERT model with the word2vec model. On this basis, the reduced performance of our model indicated that every part of our model exerts a certain effect after the removal of a certain part of the model.

5 Conclusion

The problem of multimodal detection of fake news is investigated in this study. Through the extracted OCR text, the correlation between OCR text and news text is explored, and the similarity between image statistical features and news text is calculated, thereby further extending the correlation between multimodal information. CEMM overcomes the problem of underutilization of image information (OCR text and statistical features) in the fake news detection problem. Experimental results have demonstrated the effectiveness of CEMM.

Acknowledgements. This work is supported by National Science Foundation of China No. 61702216, 61772231, and Higher Educational Science and Technology Program of Jinan City under Grant with No. 2020GXRC057, 2018GXRC002.

References

1. Allcott, H., Gentzkow, M.: Social media and fake news in the 2016 election. J. Econ. Perspectives **31**(2), 211–36 (2017)

2. Chen, J., Wu, Z., Yang, Z., Xie, H., Wang, F.L., Liu, W.: Multimodal fusion network with latent topic memory for rumor detection. In: 2021 IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6. IEEE (2021)
3. Chen, Q., Zhu, X., Ling, Z.H., Wei, S., Jiang, H., Inkpen, D.: Enhanced lstm for natural language inference. In: Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), vol. 1, pp. 1657–1668 (2017)
4. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
5. Hameleers, M., Powell, T.E., Van Der Meer, T.G., Bos, L.: A picture paints a thousand lies? the effects and mechanisms of multimodal disinformation and rebuttals disseminated via social media. Polit. Commun. **37**(2), 281–301 (2020)
6. Jin, Z., Cao, J., Zhang, Y., Zhou, J., Qi, T.: Novel visual and statistical image features for microblogs news verification. IEEE Trans. Multimed. **19**(3), 598–608 (2017)
7. Jin, Z., Cao, J., Guo, H., Zhang, Y., Luo, J.: Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In: Proceedings of the 25th ACM international conference on Multimedia, pp. 795–816 (2017)
8. Khattar, D., Goud, J.S., Gupta, M., Varma, V.: Mvae: Multimodal variational autoencoder for fake news detection. In: The World Wide Web Conference on, pp. 2915–2921 (2019)
9. Mikolov, T., Chen, K., Corrado, G.S., Dean, J.: Efficient estimation of word representations in vector space. In: ICLR (Workshop Poster) (2013)
10. Parikh, A.P., Tackstrom, O., Das, D., Uszkoreit, J.: A decomposable attention model for natural language inference. In: Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing, pp. 2249–2255 (2016)
11. Qi, P., Cao, J., Yang, T., Guo, J., Li, J.: Exploiting multi-domain visual information for fake news detection. IEEE (2019)
12. Qian, S., Wang, J., Hu, J., Fang, Q., Xu, C.: Hierarchical multi-modal contextual attention network for fake news detection. In: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 153–162 (2021)
13. Singhal, S., Kabra, A., Sharma, M., Shah, R.R., Chakraborty, T., Kumaraguru, P.: Spotfake+: a multimodal framework for fake news detection via transfer learning (student abstract). In: Proceedings of the AAAI Conference on Artificial Intelligence, vol. 34, pp. 13915–13916 (2020)
14. Singhal, S., Shah, R.R., Chakraborty, T., Kumaraguru, P., Satoh, S.: Spotfake: a multi-modal framework for fake news detection. In: 2019 IEEE fifth international conference on multimedia big data (BigMM), pp. 39–47. IEEE (2019)
15. Soranamageswari, M., Meena, C.: Statistical feature extraction for classification of image spam using artificial neural networks. In: 2010 Second International Conference on Machine Learning and Computing, pp. 101–105 (2010)
16. Xue, J., Wang, Y., Tian, Y., Li, Y., Shi, L., Wei, L.: Detecting fake news by exploring the consistency of multimodal data. Inf. Process. Manage. **58**(5), 102610 (2021)
17. Zhou, X., Wu, J., Zafarani, R.: [... formula...]: Similarity-aware multi-modal fake news detection. Advances in Knowledge Discovery and Data Mining 12085, 354 (2020)